

A Controlled Vocabulary for Biblical Studies

Sean Boisen, Logos CSO, <[myfirstname]@logos.com>

BibleTech:2010, San Jose, CA March 27 2010

Presentation URL: <http://semanticbible.org/other/talks/2010/BibleTech-LCV/BibleTech-LCV.html>

Space bar for next slide

Outline

- Semantics: an Overview
- The Logos Controlled Vocabulary (LCV)?
- Use Cases
- References for Concepts
- Concept "Importance"
- User Contribution (Logos Topics)

What Does "Semantics" Mean?

Data that has:

- Identity
 - Unambiguous
 - Addressable
- Utility
- Relationships

Semantics and Linked Data

Compare TimBL's rules for [Linked Data](#):

1. Use URIs to identify things (= Identity)
 - Use HTTP URIs so people can look things up
2. Provide useful information in a standard format when someone references a URI (= Utility)
3. Include links to other URIs (= Relationships)

Semantic Organization at Logos

Logos ... organizing the world's Biblical information

- the Bible Knowledgebase: people, places, and things described in the Biblical text
- the Logos Controlled Vocabulary
- Bob Pritchett's talk at O'Reilly Tools of Change for Publishing Conference 2010: [Domain-Specific Tools to Add Value to E-Books](#)

Other Linked-Data Semantics in Logos

Data	Id	Utility	Links
Word in text	Word number	Morphology, gloss	Lemma, cross-lingual association, Louw-Nida domain
Lemma	Lemma + homograph #	Definition, pronunciation, etymology	Other lexicons, word study
Bible Reference	Normalized Reference	Bible text	Other versions, commentaries, ...
Article	Article Id	Article content	"See also" links, parallel content

- ... and other examples
- Content and links not always formally represented

The Logos Controlled Vocabulary (LCV)

- LCV = a *controlled vocabulary* for Biblical studies
 - **controlled vocabulary**: an organized system for labeling content using words and phrases, to improve search and navigation
 - Mitigates problems of ambiguity, homographs, synonyms, variation in spelling and other kinds of representation
- Organizes content from Bible dictionaries, encyclopedias, and other subject-oriented secondary resources, regardless of how the publisher expressed it
- Provides *semantic organization* of concepts
 - Modeling the domain rather than annotating texts
 - Supporting the end-user rather than the philosopher

LCV Example: Prisca

- Example: %Prisca_Person/"Prisca"

- Utility:

Type: Person	Name: True
Preferred label: "Prisca"	Extra-biblical: False
Alternate labels: "Priscilla"	Compound: False

- Relationships:

BK Entities: #Prisca. I
Articles: Anchor/PRISCAPERSON, TYNBIBDCT/L4559, ISBE/PRISCA, ...
Logos Topics: http://topics.logos.com/Prisca
Wikipedia: Priscilla and Aquila

LCV Example: Deceit

- Example: %deceit/"Deceit"

- Utility:

Type: none	Name: False
Preferred label: "Deceit"	Extra-biblical: False
Alternate labels: "Deception", "Deceitful", "Deceive"	Compound: False

- Relationships:

Articles: ISBE/DECEIT, NBD/R494, ...

Logos Topics: http://topics.logos.com/Deceit

LCV Development

- Ambiguity: "communion" as Lord's supper, or as fellowship
- Homographs: "bear" as carrying, versus an animal, versus a constellation
- Variation: "Pilate" vs. "Pontius Pilate" vs. "Pilate, Pontius"

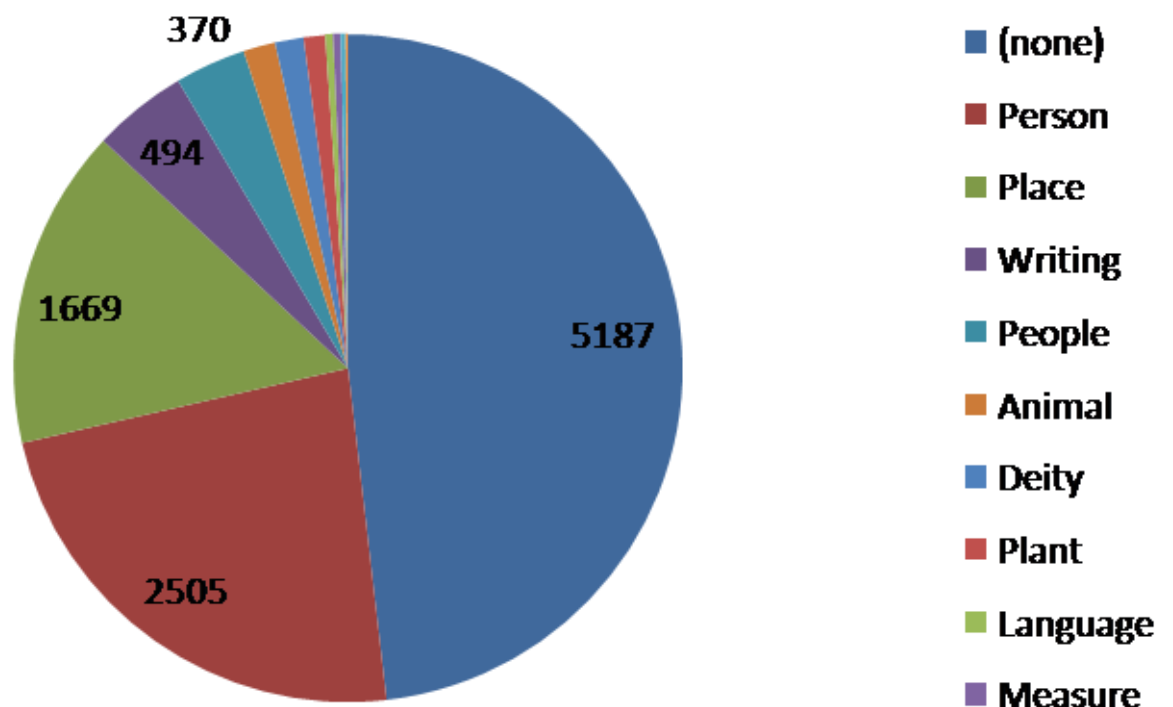
LCV Development

- Initial version developed by merging content from 7 Bible dictionaries
 - Extract headwords
 - Align automatically (but conservatively), merging when possible
 - Review manually
- Total initial terms: > 40k
- After automatic merging: ~23k
- Current inventory: ~11k concepts

LCV Development Continues

- Merging additional resources suggests new concepts:
 - Archaeological Dictionary of the Holy Land: 90/547 (16%)
 - Mostly very specific locations like %EinSamiya_Place
 - Nelson's Illustrated Bible Dictionary: 200/4833 (4%)
 - Harper's Bible Dictionary: 81/2962 (3%)
 - May be some overlap among these, particular for archaic KJV-based concepts
- Significant holes still remain:
 - Individuals from church history
 - More specialized theological concepts

LCV Types



Some concepts are still general: %Simon_Person refers to content about *multiple* Biblical people with that name (Compound:True)

Use Cases

- Improved "topic" search:
 - Link to the same concept regardless of how originally labeled
 - Provide consistent semantics for content
 - Suggest alternate concepts for the same term
- Support exploration by suggesting concepts for a passage

From Concepts to References

- Content about Biblical subjects typically includes references
- Very useful to provide the reverse index: given a passage, what concepts are associated with it?
- Estimate this by data mining and aggregating counts across all content and concepts:
 - Each article "votes" on the most likely references for the concept
 - Each concept "votes" on the most likely concepts for a reference
- Ideally, these estimates should get *better* as more content is added

PRISCA [prɪsˈkə] (*Gk. Priska*), **PRISCILLA** [prɪ sɪlˈə] (*Priskilla*). A "tentmaker" (possibly leatherworker) known both as Prisca ([Rom. 16:3](#); [1 Cor. 16:19](#); [2 Tim. 4:19](#)) and Priscilla (a diminutive form, used in Acts), the wife of Aquila, also a "tentmaker." It has been suggested that Prisca was connected with the *gens Prisca*, a noble Roman family. She and Aquila left Rome because of Claudius' expulsion of the Jews from Rome and went to Corinth, where they hosted and worked with Paul ([Acts 18:2–3](#)). From Corinth they went with Paul to Ephesus where they explained "the way of God more accurately" to Apollos ([vv. 18–26](#)). After the death of Claudius they went back to Rome, where she and Aquila were the first to receive Paul's greetings from Corinth ([Rom. 16:3](#)).

Reference Aggregation for %Prisca Content

Reference	Count	Reference	Count
Ac 18:26	10	2 Ti 4:19	7
I Co 16:19	4	Ac 18:18	4
Ro 16:3	3	Ac 18:2	3
Ro 16:3-5	3	Ro 16:4	2
Ac 18:25	2	Ac 18:2-3	2
I Co 1:12	2	I Co 16:12	2
Ac 18:11	2	Lk 8:3	1

Reference Aggregation for %Gabriel

Reference	Count	Reference	Count	Reference	Count
Lk 1:11-20	6	Tob 12:15	2	1 Th 4:16	1
Re 8:2	5	Ge 37:15	2	Da 10:5	1
Da 8:16	5	2 Ch 32:21	2	Lk 1:11	1
Da 9:21	4	Heb 7:25	1	Re 9:1-11	1
Lk 1:26-38	4	Da 10:21	1	Ex 3:2	1
Da 12:1	3	Lk 20:7	1	Da 8:18	1
Lk 1:19	3	Lk 1:31	1	Ro 14:10	1
Da 9:22	2	Da 9:21-22	1	Da 10:18	1
Lk 1:26-33	2	Da 9:21-27	1	Da 10:13	1
Dt 34:6	2	Re 8:6	1	Da 8:15-26	1

Challenges with Reference Counts

- How to combine counts
 - across overlapping references
 - for reference intersection with passage queries
- References might count less
 - if an article is longer
 - if the reference occurs later in an article
- Longer references are often more weakly "about" a concept
 - e.g. Is 40-66 for %Isaiah_Person
- We only know *that* a reference occurs, not why, or what it means
- We also manually provide "key references" for a small number of concepts

Reference-Based Retrieval: Acts 18.26

Missionaries, All Christians Should Be As



Reference-Based Retrieval: Luke 1.11-13



Concept "Importance"

- Concept content can be both *broad* (many sources) and *deep* (many words or references)
- Concepts can be popular (based on user activity)
- Concepts can be significant (by some objective standard)

Logos Topics

- Selected attributes for all concepts on the web
 - Preferred and alternate terms
 - Key passages
 - Wiki format (but more like [dmoz](#) than Wikipedia)
- Goal: engage user community in enriching data
 - Reading lists
 - Web links

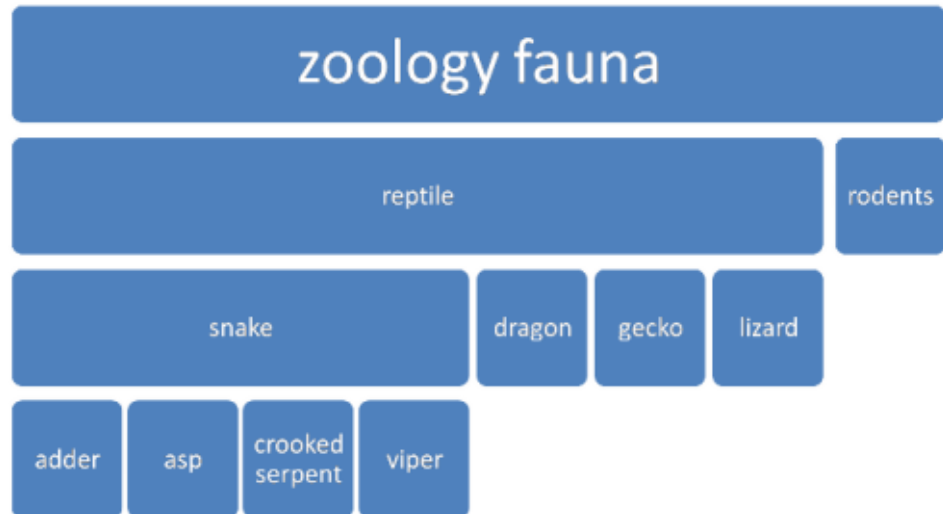
User Contributions to Logos Topics

Site was launched in August 2009, with 10779 pages (one per concept). Since that time:

- Reading lists have been added to 118 pages, including:
 - [Leviathan](#)
 - [Steward](#)
- 259 additional pages have been created, including:
 - [Book of Josippon](#)
- > 1400 other page edits, including:
 - [Aaron's Rod](#)

Future Directions

- Definitions
- Hierarchy to support broadening/narrowing search
- Incorporate user contributions
- Linkage to WordNet and similar resources



Conclusions

- Described a controlled vocabulary for Biblical studies
 - Supports search and discovery for subject-oriented content
 - Shipping now, with aspects on the web
- A growing body of linked data for Biblical studies

Resources and References

- Berners-Lee, T (2006). [Design Issues: Linked Data](#). Technical note.
- Berners-Lee, T (2009). [The Next Web](#). Presented February 2009 at TED2009.
- Boisen, S (2007). [Deploying Semantic Technologies for Digital Publishing: A Case Study from Logos Bible Software](#). Presented May 24, 2007 at the Semantic Technology Conference.
- Pritchett, B (2010). [Domain-Specific Tools to Add Value to E-Books](#). Presentation at O'Reilly Tools of Change for Publishing Conference 2010.

About This Presentation

- This is an XHTML document, structured and presented with [HTML Slidy](#). You can use your browser to view source and see how it works.
- Hyperlinks courtesy of [Logos Bible Software's RefTagger](#)
- The Logos editorial staff who have contributed to the Logos Controlled Vocabulary include Elizabeth Borries and Johnny Cisneros.